

E-SERVICES TRANSLATION PORTAL SYSTEMBACKGROUND OF THE INVENTIONFIELD OF THE INVENTION

5 The present invention is directed to a translation
portal for at least partially translating data and
facilitating the completion of the translation process.
More in particular, the present invention is directed to a
system for translating data which includes a source of data
10 to be translated, a network connected to the source of
data, a translation source connected to the network, and a
portal system connected to the network for retrieving the
data to be translated, and at least partially translating
that data. Still further, the portal system includes a
system for transmitting the at least partially translated
15 data to the translation source for completing the
translation of the data. Still further, the portal system
includes a file storage memory for storing the data to be
translated and a database for storing therein previously
entered data to be translated associated with prior
20 translations thereof.

PRIOR ART

5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000

Currently, there exists individual translation memory tools for use on a translator's workstation. Such translation aids analyze documents on a word-by-word basis, treating each new document as a totally new project. Such systems suffer from reduced efficiency in that they fail to take into account redundancies found in a document, similarities of one document to a previously translated document, and provide no means to enable team collaboration or involve the customer in the translation process.

There is therefore a need to provide a centralized translation database from which prior translations can be utilized to at least partially translated new documents to be translated. There is further a need to involve the translation customer in an iterative process, with intermediate approvals of translation work being obtained as a translation project progresses.

In an era where businesses are able to take advantage of a world wide marketplace utilizing a global computer network (Internet) it is important that such businesses are able to easily solicit business in multiple languages. Therefore, many businesses desire to have their Web pages translated into multiple languages, so that they are able to solicit business in many different markets. Therefore, a system which can upload a Web page and duplicate it in multiple languages is highly desirable. Further, as much

multiple languages is highly desirable. Further, as much of the language of any one Web page is similar to that of other Web pages, it is further desirable to make use of the translations of previously translated Web pages to aid in the translation of other Web pages. By such an arrangement, the present invention reduces the workload of translators, whether it is translation of Web pages or other documents.

SUMMARY OF THE INVENTION

5 A system for translating data is provided. The translating system includes a source of data to be translated, and a network connected to the source of data. The system further includes a translation source connected to the network and a portal system connected to the network for retrieving the data to be translated. The portal system includes a system for at least partially translating the data and a system for transmitting the at least partially translated data to the translation source for completing the translation of the data.

10 From another aspect, a system for translating data transmitted electronically through a global computer network is provided. The system includes at least one user terminal coupled to the global computer network for transmitting and receiving user related data therethrough. The system also includes at least one vendor terminal coupled to the global computer network for transmitting and receiving vendor related data therethrough. The system includes a first computer coupled to the global computer network for exchanging user related data with the at least one user terminal and vendor related data with the at least one vendor terminal through the global computer network. A file storage memory is provided which is coupled to the first computer for storing user related current document data representing documents to be transmitted from the at

least one user terminal. A second computer is provided that is coupled to the first computer and the file storage memory for at least partially translating the current document data. The system also includes a database of previously entered document data and associated translated document data coupled to the second computer. The second computer compares the current document data with the previously entered document data to match the current document data with corresponding translated document data in the database to form the at least partial translation of the current document data. The at least partial translation is sent to the first computer and transmitted therefrom to the at least one vendor terminal for completing translation of the current document data and uploading the completed translation to the first computer. The first computer includes a system for transmitting the completed translation to the at least one user terminal and the second computer. The second computer includes a system for adding the current document data and the completed translation to the database.

5

005527
10
15

20

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is an architectural overview diagram of the present invention;

FIG. 2 is a block diagram illustrating the present invention;

FIG. 3 is a block diagram illustrating the translation and management architecture of the present invention;

FIG. 4 is an illustration of the translation memory architecture of the present invention;

FIG. 5 is a flow chart representing the translation process of the present invention;

FIG. 6 is a flow chart illustrating the order process of the present invention;

FIG. 7 is a flow diagram illustrating the glossary translation process of the present invention;

FIG. 8 is a flow chart illustrating the document translation process of the present invention;

FIG. 9 is a flow chart illustrating the redundant translation process of the present invention;

FIG. 10 is an illustration of the order of submission process of the present invention;

FIG. 11 is an illustration of the translation order assignment process of the present invention;

FIG. 12 is an illustration of the order completion flow of the present invention;

FIG. 13 is a flow chart illustrating the glossary creation process; and

FIG. 14 is an illustration of the text comparison process of the present invention.

005100-001500

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

The present invention is directed to a language translation system implemented as a portal on a global computer/communications network and combines advanced translation technologies with human translators for carrying out cost-efficient multilingual electronic commerce initiatives.

In FIGS. 1-4, system 1000 is illustrated. System 1000 provides a language portal implemented on the Internet. A plurality of customers 2 access the translation server network 8 through the global computer network (Internet) 6. As will be described further in following paragraphs, documents submitted by customers 2 to the translation server network 8 are translated at least partially automatically by the localization engines 18 utilizing the multilingual translation database 20, by offline translators 4 and/or online translators 26, through an intranet network.

Looking at the operation of system 1000, one or more users 2, using a Web browser or like type software access the Web server 10 through the global computer network 6. The E-services Translation Portal (ETP) 10 allows users 2 to establish translation projects and store ETP and flow data in the flow database 14 for establishing a workflow for the customer's documents to be translated, also

providing reports to the administrator 22 and the flow designer 24. As the number of customers increases, additional servers 16 are brought online. Customer documents to be translated are stored in the file system memory 12. The documents stored in file system storage 12 are subsequently processed by localization engines 18 wherein the documents are identified as to file type, i.e. HTML, GIF, any one of a multiplicity of word processor, text, or spreadsheet file types. Next, the file is separated into translatable and non-translatable strings and stored in the translation database 20. Once the translatable strings of the file have been identified, they are then compared with previously translated strings to find similar strings which have already been translated. If a string matches more than one previously translated string, the customer's string to be translated at this time will be matched with the prior translation which has the highest frequency of use, or from the latest or user specified single or multiple previous translation versions. The partially translated document is also stored in the translation database 20.

The translation of the partially translated document is then completed by an online translator 26, or an offline translator 4, in accordance with the workflow design established for the particular customer 2. The workflow

design may specify a particular translator 4, 26 that the customer 2 has engaged. The vendor (translator) assigns a price per unit for its work. This unit can be based on the number of words to be translated, a number of hours, etc.

5 The customer selects a vendor based on a number of criteria including, but not limited to, resumes, references, customer ratings, volume, location, expertise and the like, which information is transferred to the customer 2 by the server 10.

10 Customers submit projects through the Internet to the translation server network, where the Web server/ETP 10 and database 14 are utilized to provide a measurement of the project. The size of the project is automatically estimated and broken down by some number of units. Then, based on the number of units and the vendor's cost per unit, a quote is provided to the customer through the server 10, which can also include a schedule and incremental costs. Alternately, the project may be entered for bid, where the customer submits its project to a
15 selected list of vendors from which quotes are received.
20

After the customer receives and approves a quote, the customer transmits a contract to the vendor to perform the work. The selected vendor then executes the contract electronically to accept the project. The fact that the
25 vendor receives the project from the translation server

network 8, partially translated, reduces the costs of translation services. The customer, using their connection to the translation server network 8 through the Internet, can track the progress of the project, conduct interim reviews, provide interim approvals and, through the use of log files, has access to an audit trail.

Turning now to FIG. 2, such shows system 1000 from another perspective. Access to Web server/ETP 10 through the network 6 is accomplished by customers utilizing browsers for submitting translation jobs as part of HTML pages 28. Vendors access the documents to be translated through browser based tools 30, 32 and 34. Web server/ETP 10 communicates with the multilingual application server 38, the server 10 receiving the files and transmitting the files to the customer, while the application server 38 creates the projects, creates orders, checks for the translatable strings and stores them in the translation database 20. The application server 38 embodies servers 16, database 14, localization engines 18, and the file system 12 of FIG. 1. The functions of servers 10 and 38 can be implemented on a single computer or a network of computers. Multilingual application server 38 communicates with the multilingual translation memory database 20 through the intranet 64, the network which is internal to the translation server network 8, shown in FIG. 1. The

translatable strings which are stored in translation database 20 may also be processed by third party linguistic analysis utilities 48 which may access the database. The translation memory database 20 is also accessible by the online translators 26 through intranet 64.

Turning to FIG. 3, such shows the software architecture for accomplishing the translation of input documents. The files stored in the file system storage 12 may be in any of a plurality of different text documents or graphics formats 52a, 52b, 52c. The localization engines 18 utilize utilities 54a to interpret those differently formatted documents, and utilizes on-line manager and editor routines 54b for processing those documents. The processing of those documents is done in cooperation with an open system multilingual library 56 and the localization engines 18 communicate with the translation database 20 through an application program interface (API) 66 utilizing an open protocol.

Referring to FIG. 4, such illustrates the translation memory architecture. Translation database 20 stores multiple projects 100, 108 which each include respective strings 102, 104, 106 and 110, 112, 114, 116 to be translated. In the illustration, menus for use in a computer program are illustrated. The translated projects 130 and 138 are also stored in memory with the respective

translated (French) strings 132, 134, 136 and 140, 142, 144, 146. A database text table 120 is provided which includes pointers for finding the text in translation database 20. Each entry 120a-120n in the table 120 includes an ID which refers to the particular project, the language of the text, the text itself and a reference count, which indicates the number of different projects utilizing that same text. Thus, the text "open file" 102 and 112 is referred to as a single entry 120a in the table 120. Similarly, the translated French text "ouvrir fichier" 132 and 142 are similarly referred to in the single database entry 120b.

Where text incorporates a translated term in some other project n, such is considered a subset of the preceding entry. Thus, the text "open file" is considered a subset of the text "open file to load data" and thus the entry 120n will include a reference pointing to the Table entry 120a. By that arrangement, a portion of the string "open file to load data" can be obtained from the previously translated string "open file". Thus, the translation process is simplified and the database overhead is reduced. By maintenance of the translation database 20, the automated translation system becomes more robust the more it is used, as every string which has been previously translated need not be duplicated by human translators.

The translation database 20 can therefore be shared by multiple customers and vendors through the Internet. Localization engines 18 search the database for past translations. Translations which match, or just partially match are identified and added to the file that is provided to the translation vendor 4, 26.

The advantages provided by the system can be clearly seen when one considers a class of customers who would utilize the system. For instance, Web site owners who wish to duplicate their Web sites in multiple languages are customers who would benefit from system 1000. The text found in most Web sites is between 50% and 60% redundant. Therefore, at least 50% to 60% of the time and money required for translation would be saved, as the redundant text strings found in a given HTML file of a web site would be translated automatically, and not require the services of the vendor translator. Thus, over time, users of system 1000 would increasingly benefit by its expanded knowledge database.

An example of a translation process workflow which would be established by the flow designer 24 is shown in FIG. 5. The process starts at 200 and next goes to the decision block 202, wherein it is determined whether a glossary is needed. A glossary is a collection of long phrases found in a document. In some cases, the customer

may already have a glossary that is uploaded, in other cases it must be created. Glossaries are particularly useful where there are many translators working on a project. One translator translates the glossary and then after review and approval by the client, or a reviewer contracted by the client, it can be distributed to all the translators. The use of glossaries ensures that common terms are translated consistently across all files of a customer's single project or multiple projects. The instant invention provides for creation of both default and custom glossaries. When a default glossary is created, the strings are extracted automatically, and are therefore project specific. Custom glossaries can be created to standardize terminology across projects, corporate divisions, or entire corporations. The user can create as many glossaries as desired. The glossaries are not used to automatically translate a project, but are given to all translators working on a client's project or projects, to give all the translators the same verbiage for use within their respective translations. The use of the common verbiage of the glossary keeps the consistency at a high level.

If a glossary is needed, the flow passes to block 204 where the glossary generation process is initiated. From block 204, the flow passes to decision block 206 to

determine whether there is redundant data. If there is redundant data, the flow passes to block 208 wherein redundant processing is initiated. Thus, the reference count for the text strings are scanned and only one string of a redundant group need be translated. From block 208, the actual document translation process is initiated in block 210 and the translated document is generated in block 212. From block 212, the flow passes to block 214, wherein alterations to the text document are made, such as changes to the font size and format of the document. From block 214, the flow passes to block 216, wherein the customer inspects the document. From there, the flow passes to block 218, wherein the customer's final approval is received. From block 218, the flow passes to decision block 220 where it is determined whether the document has been finally reviewed. If it has, then the flow passes to block 226 wherein the customer is billed. If the final review has not been completed, then the flow passes to block 222 wherein the document is checked, and if problems are found, the flow passes back to block 214 to cycle through the loop again. If there are no problems, or such have been fixed, the flow passes from block 222 to block 224 wherein the document is updated, as required. From block 224, the flow passes to block 226, wherein the customer is billed. From block 226, the flow passes to

block 228, to indicate to the parent process, such as the order process shown in FIG. 6 that workflow is complete.

In FIG. 6, the order process is shown. Starting from the start block 250, the flow passes to block 252. In block 252, an account manager is assigned to the project that will be created in block 254. In block 254, the project and version (which language) is created by the customer and can include creation of the workflow design (FIG. 5). From block 254, the flow goes to decision block 256, wherein it is determined whether the project and version are created properly. If they are, the flow passes to block 260, otherwise, the flow passes to block 258, wherein the assigned account manager checks and fixes the project and/or version. From block 258, the flow passes to block 260. In block 260, the application programming interface 66 and the appropriate utility 54A based on the file format of the document are called to extract the translatable strings. From block 260, the flow passes to decision block 262, wherein it is determined whether the strings have been completely extracted from the document. If the strings have been completely extracted, the flow passes to block 266. If they have not, flow passes to block 264, wherein the account manager checks the document and fixes any problems therewith that are preventing the complete extraction of the translatable strings. From

block 264, the flow passes to block 266, wherein it is determined how many target languages are to be provided for the client, so the proper number of copies of the original file can be made. From block 266, the flow passes to decision block 268 wherein it is determined whether there is a need for a glossary. If a glossary is not required, flow passes to block 278. If, on the other hand, a glossary is required, the flow passes to decision block 270 wherein it is determined whether the customer has provided a glossary. If the customer has provided a glossary, the flow passes to block 274 where the glossary is imported and then flow passes to block 278. However, if no glossary is provided, the flow passes to block 272 wherein a glossary is created. Subsequent to creation of the glossary, the flow passes to block 276 wherein the glossary is reviewed by the client. From block 276, the flow passes to block 278 wherein the client selects the translator and reviewer (usually two different persons) for each target file. Subsequent to selection of the vendors in block 278, the flow passes to block 280, wherein the customer confirms the order. From block 280, the flow passes to decision block 284, wherein it is determined whether the order has been confirmed. If it has not, the flow passes to block 282 wherein the account manager fixes the problem in obtaining an order confirmation. In some cases, a client may have

desired translation of their project into a particular number of target languages, for instance ten, but the final cost for that quantity of translations may result in a quoted cost which exceeds the client's budget. Therefore, the client may then have to scale back the extent of their order or find new vendors offering lower costs. Once the problems are fixed in block 282, the flow passes back to block 280, to repeat the cycle in obtaining order confirmation. After the order has been confirmed, the flow passes from block 284 to block 286 wherein the translation process is initiated. The translation process is the flow diagram of FIG. 5, thus the process of FIG. 6 is the parent process of that disclosed in FIG. 5. Therefore, the notification provided in block 228 of FIG. 5, is provided in block 286 of FIG. 6. From block 286, the process ends at block 288.

FIG. 7 is a flow diagram of the glossary translation process which is initiated in block 204 of FIG. 5. From the start block 300, the flow passes to block 302, wherein a translator 4 downloads the glossary from the Web server 10. The selected vendor 4 is sent an E-mail message informing them that the glossary is ready to be downloaded and provides a hyperlink to the Web page identifying the site from which the glossary may be downloaded via Web based file transfer mechanisms including but not limited to

File Transfer Protocol (FTP). From block 302, the flow passes to block 304, wherein the vendor provides the translation of the glossary into the desired target language or languages. From block 304, the flow passes to block 306, wherein the translated glossary is provided to the customer for the customer's review. To insure that the customer agrees with the translation, a loop is provided which begins with decision block 308 where it is determined whether or not the customer has approved the translated glossary. If the customer has not approved it, flow passes to block 310, wherein the vendor fixes the problems identified by the customer. From block 310, the flow passes back to block 306, to permit the customer to review the corrections made by the vendor. Once the customer has approved the glossary, the flow passes to block 312 wherein the translated glossary is attached (identified with) the document to be translated. From block 312, the flow passes to block 314, wherein the parent process, shown in FIG. 5, is notified that the glossary translation process has been completed. From block 314, the process terminates at block 316.

Block 210 of FIG. 5 initiates the document translation process. The flow chart for that process is illustrated in FIG. 8. From start block 320 in FIG. 8, the flow passes to block 322 wherein the vendor is sent an E-mail message

describing the Web site location of the documents to be translated, so that the vendor can inspect the documents. Flow then passes to block 324, wherein the translator downloads the documents for offline translation. The flow then passes to block 326, wherein the translator performs the translation of the documents. Subsequently, upon completion of the translation, flow passes to block 328, wherein the translated documents are forwarded (uploaded) for review. The reviewer is sent an E-mail indicating that the translation is ready for review. Flow then passes to block 330 wherein the reviewer downloads the translated documents. Subsequent to downloading the documents, the flow passes to block 332 wherein the reviewer performs a review of all of the translated documents and makes any corrections necessary. From block 332, the flow passes to block 334, wherein the reviewed documents, and especially corrections, are checked. Such secondary review, may be carried out by a second reviewer. From block 334, the flow then passes to block 336 wherein the reviewed translated documents are uploaded to the Web server 10. The flow then passes to block 338, wherein the parent process, block 210 of FIG. 5, is notified that the translation process is complete. The process then ends at stop block 340.

Turning now to FIG. 9, such shows the redundant string translation process which is initiated by block 208 of FIG.

5. From the start block 350, flow passes to block 352 wherein the redundant strings are downloaded by the translator. From block 352, the flow passes to block 354, wherein the translator translates the strings which have been identified as having multiple re-occurrences in the document. From block 354, the block passes to block 356, wherein a reviewer, usually a separate individual, reviews the translated strings. From block 356, flow passes to block 358, wherein the parent process is notified that the redundant string processing has been completed, such notification being provided in block 208 of FIG. 5. From block 358, the process terminates at block 360.

Referring now to FIG. 10, such shows the order submission process from another perspective. The customer 2 utilizing a Web browser submits an order to the E-services translation portal embodied in the Web server 10 through the Internet 6. The server 10 stores the documents in a file system disposed within the memory 15. Memory 15 includes file system 12 and translation database 20 shown in FIG. 1. Server 10 also launches the workflow extraction process to extract the translatable strings which are stored in the translation memory database of memory 15, which is carried out by the workflow process 80. The workflow process 80 also includes the reuse of prior translations, where a search is made between the

translatable strings of a current project and strings which have previously been processed, to provide a translation when a match is found.

In FIG. 11, the translation order assignment flow is shown. The workflow process 80 creates a project file and stores that in the file system of memory 15. The workflow process 80 then sends an E-mail 82 to the translator 4. The E-mail contains information about the order, the download location, a hyperlink to the ETP page having a link to the Internet site for downloading the file(s) via Web based file transfer mechanisms including but not limited to File Transfer Protocol (FTP). The translator 4 utilizing the Web browser browses the ETP Web page containing the Web download link to the project file (e.g. via FTP) by utilizing the hyperlink contained in the E-mail 82. Using the browser, the translator 4 downloads the project file from the Web download site and then invokes the local translation software and proceeds to then perform the translation task.

Upon completion of the translation task, the translator then uploads the translated document, which process is illustrated in FIG. 12. The translator 4 utilizing a Web browser uploads the translated project through the Internet 6 to the ETP server 10. The server 10 saves the project to the file system in memory 15 and

notifies the workflow process 80 that the files are ready. The workflow process 80 handles the flow from reviewers to accept or reject the translation. If the reviewer accepts the project files then the workflow process 80 will store the translated document in the translation memory database, within memory 15. Finally, the translated files are provided to the customer. In order to complete that process, the workflow process 80 sends an E-mail to the customer which contains information about the order, the download location for the translated documents and a hyperlink to the ETP page with a Web download link to the project file (translated documents) via Web based file transfer mechanisms including but not limited to File Transfer Protocol (FTP). A customer, utilizing their browser, employs the hyperlink to access the Web download link to the project, and then utilizing that link, downloads the translated file.

The process for creating a default glossary, block 272 of FIG. 6, is shown in FIG. 13. In FIG. 13, the flow starts with a decision block 400 wherein it is determined whether a user interface and thus texts included in the user interface exists or whether there is seed data. Seed data is a list of terms which come with initial software installation or are related to and are usually peculiar to a particular industry. If there is text embedded in a user

interface - or seed data, the flow passes to block 402 , wherein text strings are extracted from that data or user interface to create what is termed a type 1 glossary. If no user interface or seed data is provided, or subsequent to creating the type 1 glossary in block 402, the flow passes to block 404 wherein a linguistic engine, such as use of an electronic dictionary is used to try and match up long phrases and create a type 2 glossary. From block 404, the flow passes to block 406, wherein redundant strings, strings with more than one reference count and having less than five words are further identified as a type 3 glossary. From block 406, the flow passes to block 408, wherein the three types of glossaries are merged together and redundancies removed. From block 408, the block passes to block 410 where a customer glossary, if provided, is imported into the system.

Flow then passes to block 412, wherein the merged glossary is compared against the customer's glossary to find any conflicts. If conflicts are found, the flow passes to block 414, wherein the customer is requested to review the conflicts and make additions or corrections to the generated glossary. Flow, of course, would then pass back to the parent process.

As previously discussed, the translation database of the present invention provides great efficiencies in

reducing the translation work load required of vendors. As an example, FIG. 14 is provided to illustrate the comparison which is carried out utilizing the translation database. Thus, when a new project 108, having

5 translatable strings 110, 112, 114 and 116, to be translated is entered into the system, it is compared with prior projects that have been provided. The match need not be exact, so that another project having the same or similar strings may be found. The comparison may then

10 identify a prior project 100 having strings 102, 104 and 106 as being identified as being a close match. The difference between the project 100 and the new project 108 is a single string 110, thus the translation that was

15 utilized for project 100 may be utilized again for project 108, with only the string 110 remaining to be translated. By that process, the translation workload which must be contracted out is minimized. In addition to the project level matching, system 1000 performs the translatable

20 string level searching to further reduce the manual translation workload.

Although this invention has been described in conjunction with specific forms and embodiments thereof, it will be appreciated that various modifications other than those discussed above may be resorted to without departing

25 from the spirit or scope of the invention. For example,

5